

# Fuzzy logic circulation types based on the Osuchowska-Klein classification system created for Poland

Jadwiga Woyciechowska · Zbigniew Ustrnul

Received: 1 March 2010 / Accepted: 28 October 2010 / Published online: 20 November 2010  
© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** Fuzzy logic was applied as a methodology to create an objective calendar of daily atmospheric circulation based on an existing atmospheric circulation classification system for Poland called the Osuchowska-Klein method, which includes 13 circulation patterns. In this study, circulation patterns given by Osuchowska-Klein were expressed as normalized values of sea level pressure in a regular  $5^\circ \times 5^\circ$  grid over the area of interest. Cases of daily atmospheric circulation (i.e., normalized values for the same domain) observed from January 1948 to December 2007 were defined in the same manner. The pattern-matching procedure employed in the study was made objective with the application of fuzzy logic sets in a few steps as follows: (1) discrimination and grouping of grid points as well as their classification in order to clearly define each Osuchowska-Klein circulation pattern, (2) calculation of the similarity between observed normalized pressure and that of the pattern at each selected grid point using a membership function, (3) determination of the degree of fulfillment for every class of grid points, (4) determination of the degree of a case and pattern, (5) selection of the circulation type for a particular day based on the degree of fulfillment for each pattern. The largest value of the degree of fulfillment indicates the type of circulation pattern on any given day. Finally, a validation of the obtained results is discussed. This was accomplished using the chi-square test and a comparison of consistency

between the traditional and objective approach types. The results do show some differences between the subjective and objective methods, however, they are not major differences, especially when taking into account the fact that the two methods are quite different, as well as the complex nature of pressure systems.

## 1 Introduction

In recent years, the number of so-called objective circulation type classification systems has increased rapidly (Huth et al. 2008). New computing options and other novel methods have emerged in the last two decades along with large digital meteorological databases. This has produced a whole new range of classifications based on new analytical methods and concepts. On the other hand, proven manual classification systems (subjective systems) have been abandoned, in most cases, even though their practical and scientific value was significant. Unsurprisingly, some researchers decided to make existing manual circulation typologies objective (e.g., Cawley and Dorling 1996). Two examples of well-known classification systems in Europe are the Grosswetterlagen (James 2007) and the Lamb systems (Jenkinson and Collison 1977). Even less-renowned local classification systems can be made objective and there is a purpose to that. This is exactly the case with the well-known Polish manual classification system developed by Osuchowska-Klein (1975).

The main purpose of this study was to create a 24-h objective classification system based on the existing subjective Osuchowska-Klein classification system. This classification approach is used to evaluate circulation patterns over Poland and it is commonly applied in synoptic climatological studies. The calendar with Osuchowska-Klein

J. Woyciechowska  
Institute of Meteorology and Water Management, Cracow Branch,  
Cracow, Poland

Z. Ustrnul (✉)  
Department of Climatology, Jagiellonian University,  
Cracow, Poland  
e-mail: zbigniew.ustrnul@uj.edu.pl

types is available for the 1901–1998 time periods. Unfortunately, it has not been produced since 1999. This classification system has been selected for two additional reasons. First, it provides a relatively good assessment of meteorological phenomena over Poland. This is especially true of extreme weather events (Osuchowska-Klein 1978). Second, it consists of a relatively small number of patterns/types (13 patterns), which makes it possible for it to be used even with short data series.

The system's only flaw described in some publications is the fact that Type X (the 14th type—unclassified) occurs relatively frequently and its frequency of occurrence can reach several percent. This is especially true of synoptic situations, which often produce extreme air temperatures. Type X can be eliminated by making the Osuchowska-Klein classification system objective.

## 2 Data

The original Osuchowska-Klein circulation type classification system is based on 13 patterns that depend on the location of main pressure centers over Europe, which determine air mass movements over Poland (Osuchowska-Klein 1975). The following 13 classified patterns (types) and one unclassified pattern (type) have been identified: A—Western cyclonic, CB—Northwestern cyclonic, D—Southwestern cyclonic, B—Southern cyclonic, F—Southeastern cyclonic, C2D—Western anticyclonic, D2C—Southwestern and Southern anticyclonic, G—Central anticyclonic, E2C—Northwestern anticyclonic, E0—Northeastern and eastern cyclonic, E—Northeastern and eastern anticyclonic, E1—Southeastern and eastern anticyclonic, BE—Southern transitional between cyclonic and anticyclonic, X—unclassified.

European circulation patterns were initially produced by the Osuchowska-Klein method as part of this research study and then expressed as normalized values of sea level pressure across a regular  $5^\circ \times 5^\circ$  grid over an area ranging from  $45^\circ\text{W}$  to  $60^\circ\text{E}$  and from  $30^\circ\text{N}$  to  $75^\circ\text{N}$ . Observed daily atmospheric circulation (i.e., normalized values for the same domain) from January 1948 to December 2007 (based on NCEP/NCAR Reanalysis) was determined in the same way.

## 3 Methodology and results

The method applied in our study is based on case-based reasoning systems rooted in fuzzy set theory. Case-based reasoning is a method for solving problems by remembering similar previous situations and reusing information and knowledge about those situations (Hansen 2000, based on: Kolodner 1993; Leake 1996). It enables one to indicate

similarity between a compared case (the synoptic situation on a given day) and a pattern (any of the 13 patterns in the Osuchowska-Klein classification system). This type of fuzzy logic methodology was used to determine circulation types based on the Osuchowska-Klein classification system for arrays of observed situations. The approach was applied to information acquired about the level of similarity between the features of a given case and those of a set pattern. This type of information is encoded in the so-called membership function (see further part of the text). Fuzzy logic emulates the work of experts performing comparative analysis.

The main purpose of our methodology was to create an objective calendar based on the Osuchowska-Klein classification system. A fuzzy logic approach was applied to the daily atmospheric circulation classification calendar. In a classic statement of logic, an answer is either true or false. In fuzzy logic, an expression may be true to a certain degree. An “objective” classification calendar was produced by identifying the given Osuchowska-Klein circulation pattern that best matches a given case (i.e., daily sea level pressure map). This was accomplished by calculating a similarity measure (the so-called degree of fulfillment) of each given case to each given pattern. Normalization was performed in order to make the patterns and cases comparable. European circulation patterns generated via the Osuchowska-Klein method were expressed as normalized values of sea level pressure across a regular  $5^\circ \times 5^\circ$  grid over an area ranging from  $45^\circ\text{W}$  to  $60^\circ\text{E}$  and from  $30^\circ\text{N}$  to  $75^\circ\text{N}$ . In order to normalize the pressure over the selected area, an expression of the following form (Han and Kamber 2001) was used:

$$a' = \frac{a - a_{\min}}{a_{\max} - a_{\min}} (a'_{\max} - a'_{\min}) + a'_{\min}$$

where:

$a_{\max}$	Maximum pressure value over domain
$a_{\min}$	Minimum pressure value over domain
$a'_{\max}$	Maximum value of normalized pressure over domain
$a'_{\min}$	Minimum value of normalized pressure over domain
$a$	Pressure value at the given grid point
$a'$	Normalized pressure value at the given grid point

In the study, the highest normalized pressure  $a'_{\max}$  assumes a value of 1 while the lowest value of normalized pressure  $a'_{\min}$  equals 0.

Observed cases of daily atmospheric circulation (i.e., normalized values for the same domain with the same resolution as the model patterns) from January 1948 to December 2007 (based on NCEP/NCAR Reanalysis) were expressed in the same way. Circulation patterns are distinguished by their typical locations of cyclones and

anticyclones over the domain of interest. Taking into account the location of cyclones and anticyclones as well as the boundaries that separate centers of pressure (key determinants of weather), a set of grid points that clearly defines each pattern was selected. For the purpose of this study, a sub-domain ranging from 0°E to 40°E and from 40°N to 65°N was defined. The synoptic situation in the sub-domain affects the weather over Poland the most (Fig. 1). Figure 1 also shows grid point locations that clearly define one of the patterns of interest (Type D2C). Six grid point classes  $\nu$  for cyclones, anticyclones, and areas that separate pressure centers inside and outside the sub-domain were created.

The degree of fulfillment of a given case to a pattern was obtained in a few steps. At the beginning, the difference  $x_1 - x_2$  between observed normalized pressure  $x_2$  and that of the Osuchowska-Klein pattern  $x_1$  was calculated for each selected grid point of each pattern. Then, in order to obtain the similarity  $\mu$  between  $x_1$  and  $x_2$  at each selected grid point, a so-called membership function was constructed. In order to compare continuous fields such as pressure, a membership function such as the one shown in Fig. 2 (where constant  $c = 0.125$ ) was used (Hansen 2000). The membership function enables one to calculate the degree of similarity of pressure as a function of the difference  $x_1 - x_2$  between the normalized value of pressure of the pattern  $x_1$  and the normalized value of pressure of the given case  $x_2$  at each grid point of the set of locations that define the pattern. In our study,  $\mu \equiv \mu_\nu(j, k)$  is the similarity at each selected grid point estimated using the membership function, and established for ranges of differences  $x_1 - x_2$  between case  $t$  and pattern  $k$  for individual location  $j$  and class  $\nu$ . For every grid point class  $\nu$  of the pattern  $k$ , the degree of fulfillment  $D_{k\nu}$  was calculated as follows: inside the sub-domain, *most of* the selected locations should fit the pattern while at locations outside the sub-domain, *at least a few* values at selected grid points can be expected to fit the pattern. For this reason, different values of the weight  $\gamma_\nu$  for selected locations inside and outside the sub-domain were taken into consideration in our

analysis. Satisfactory results were obtained for  $\gamma_\nu = 0.4$  inside the sub-domain and for  $\gamma_\nu = 0.7$  outside the sub-domain for each circulation pattern. The degree of fulfillment  $D_{k\nu}$  (Driankov et al. 1996; Bardossy et al. 1995) was expressed as a combination of *OR* ( $F_o$ ) as well as ( $F_a$ ) functions. In fuzzy set theory, the function  $F_o$  takes the following form (Dubois and Prade 1980):

$$F_o(\mu_\nu(1, k), \mu_\nu(2, k)) = \mu_\nu(1, k) + \mu_\nu(2, k) - \mu_\nu(1, k) \mu_\nu(2, k)$$

and recursively as:

$$F_o(\mu_\nu(1, k), \dots, \mu_\nu(j, k)) = F_o(F_o(\mu_\nu(1, k), \dots, \mu_\nu(j-1, k)), \mu_\nu(j, k))$$

While the function  $F_a$  is expressed as (Bardossy and Disse 1993)

$$F_a(\mu_\nu(1, k), \dots, \mu_\nu(m, k)) = \prod_{j=1}^m \mu_\nu(j, k)$$

The degree of fulfillment  $D_{k\nu}$  was calculated using the following equation (Driankov et al. (1996)):

$$D_{k\nu} = \gamma_\nu F_o(\mu_\nu(1, k), \dots, \mu_\nu(j, k)) + (1 - \gamma_\nu) F_a(\mu_\nu(1, k), \dots, \mu_\nu(j, k))$$

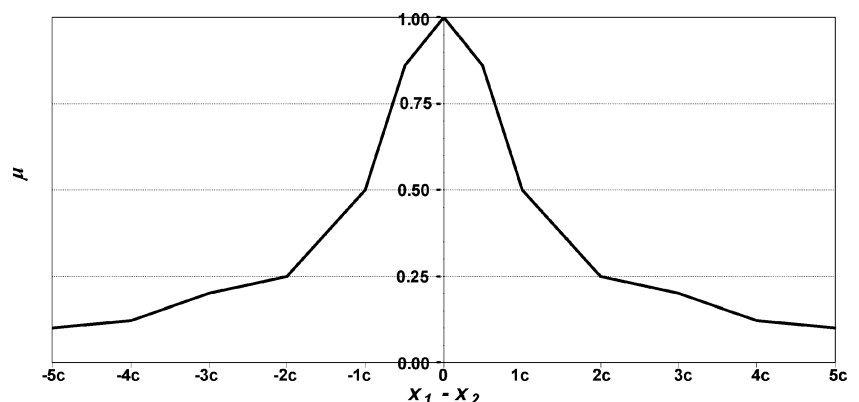
The next step was to calculate the degree of fulfillment  $D_k$  that showed similarity between the case to be classified and the pattern of interest—the degree of similarity of the circulation observed to a given type of classification (Bardossy et al. 1995). The degree of fulfillment  $D_k$  corresponding to pattern  $k$  and the case of day  $t$  is the product of the  $D_{k\nu}$  values obtained for every class  $\nu$ :

$$D_k = \prod_{\nu=1}^n D_{k\nu}$$

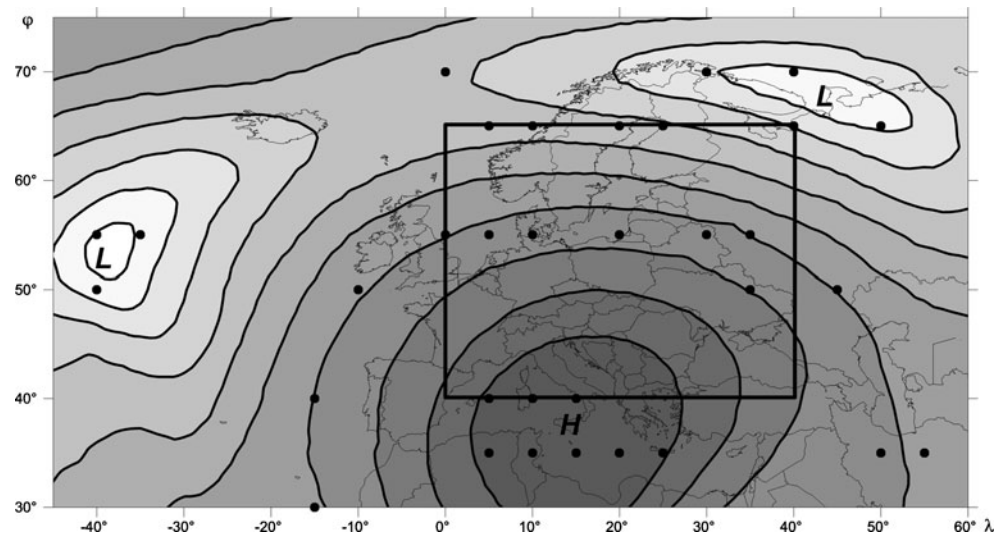
where  $n = 6$  is the number of classes.

Finally, values of  $D_k$  for each pattern and the case of day  $t$  were compared in order to select the circulation type of the observed situation  $t$ . The largest value of the degree

**Fig. 1** Membership function (Hansen 2000); see text for discussion



**Fig. 2** Example of grid point location describing the D2C type over the area of interest (domain and sub-domain)



of fulfillment  $D_k$  for a compared case indicated the type of circulation pattern on any given day. This means that the pattern  $k$  for which the degree of fulfillment  $D_k$  takes on the maximum value is selected as the circulation pattern for day  $t$ .

It must be noted that minimum similarity thresholds were not defined by the method where a given situation is assigned to one of 13 predefined Osuchowska-Klein classification types. For this very reason, the objective method does not take Type X (unclassified) into account. Another reason for this was that the days on which Type X was supposed to occur, according to the subjective method, were not compared.

#### 4 Validation

Validation turned out to be one of the most difficult tasks. The authors wanted to compare the obtained results with the Osuchowska-Klein subjective calendar. Circulation types, described using the subjective method, were compared multiple times for each day from 1951 to 1998, applying different parameters generated by the fuzzy logic method. Although the comparison itself was quite simple—the number of cases obtained for all 13 types via the objective method was compared with that obtained via the subjective method—it needs to be emphasized that this type of analysis cannot be fully reliable. The most basic reason for this is the lack of the same initial materials in the two cases. The original “subjective” calendar was based on daily synoptic maps of Europe but it is not clear whether these maps were based on several observations (mean of several observations per day) or just based on a particular observation time each given day. Thus, it was decided to use mean daily grid data in the study. As a result, the circulation type obtained for a given day is a consequence

of the situation, which took place at the main observation times. An example of this would be a comparison of a daily average of a synoptic situation and one point in time (e.g., 12 UTC) where, it turns out, the differences observed on the maps are not significant. In summary, it must be stated that the validation procedure was not carried out using the same initial data set, which would have produced similar, but not exactly the same, results from the very start.

Two different validation methods have been applied. The first was based on a statistical approach while the second consisted of a simple analysis of the contingency table supported by a simple comparison of the cases involved (comparison of both calendars and corresponding cases/days, i.e., subjective and objective ones).

##### 4.1 The first “statistical method”

The first method consisted of a chi-square goodness of fit test (which is the sum of differences between calculated and expected outcome frequencies, Wilks 2006) and a chi-square test of independence (Kendall and Stuart 1973).

The chi-square test was used to compare the subjective method and the objective method to see if they produce consistent results (distribution of types). The following formula was used:

$$\chi^2 = \sum_{n=1}^m \frac{(O_n - E_n)^2}{E_n}$$

where:

- $O_n$  Number of cases of a given type (pattern)  $n$  based on the objective method
- $E_n$  Number of cases of a given type (pattern)  $n$  based on the subjective method
- $n$  Ranges from 1 to  $m$ ;  $m=13$

The calculated chi-square value is markedly larger than the threshold value, which is 5.226 for the confidence level  $p=0.95$  and 12° of freedom. In light of this result, the objective method appears to be erroneous (not consistent). In the case of our data, this test does not seem to be sufficient. For this very reason, another test was then used, the chi-square test for independence. The formula shown below (Kendall and Stuart 1973):

$$\chi^2 = \sum_i \sum_j \frac{n_{ij}^2}{n_{i.} n_{.j}} - 1$$

where:

$n_{ij}$  Frequencies  
 $n_{i.}$  Row totals  
 $n_{.j}$  Column totals

was used to calculate the  $\chi^2$  value ( $\chi^2=22,022$ ), which exceeds even the largest threshold values many times over, given the number of degrees of freedom. The very large chi-square value indicates a strong relationship between the parameters (types) of both distributions. It was validated by the value of Pearson's coefficient of contingency ( $P=0.751$ ; Kendall and Stuart 1973), as well as by the value of a chi-square alternative (Kendall and Stuart 1973) called Tschuprow's  $T$  ( $T=0.328$ ).

In spite of unambiguous test results produced by both methods (i.e., subjective and objective), which show a highly consistent distribution of circulation types, still get the feeling that results are still somehow insufficient.

From our point of view, considering the nature of circulation phenomena, chi-square does not show which method is better. It only shows if the results produced by the two methods are in agreement. An analysis of the contingency table, which has been discussed to a significant

extent (see below), shows which classification types are “confused” most often. An analysis of specific cases shows that the identification and assignment of a classification type to a given synoptic situation is neither easy nor unambiguous.

It must be noted that when certain cases (synoptic situations, specific days) are compared directly, the objective method appears to rely more on the boundaries of pressure systems than on the locations of pressure system centers. This is especially true of Central Europe, including Poland, where centers of both high as well as low-pressure systems rarely linger over the region.

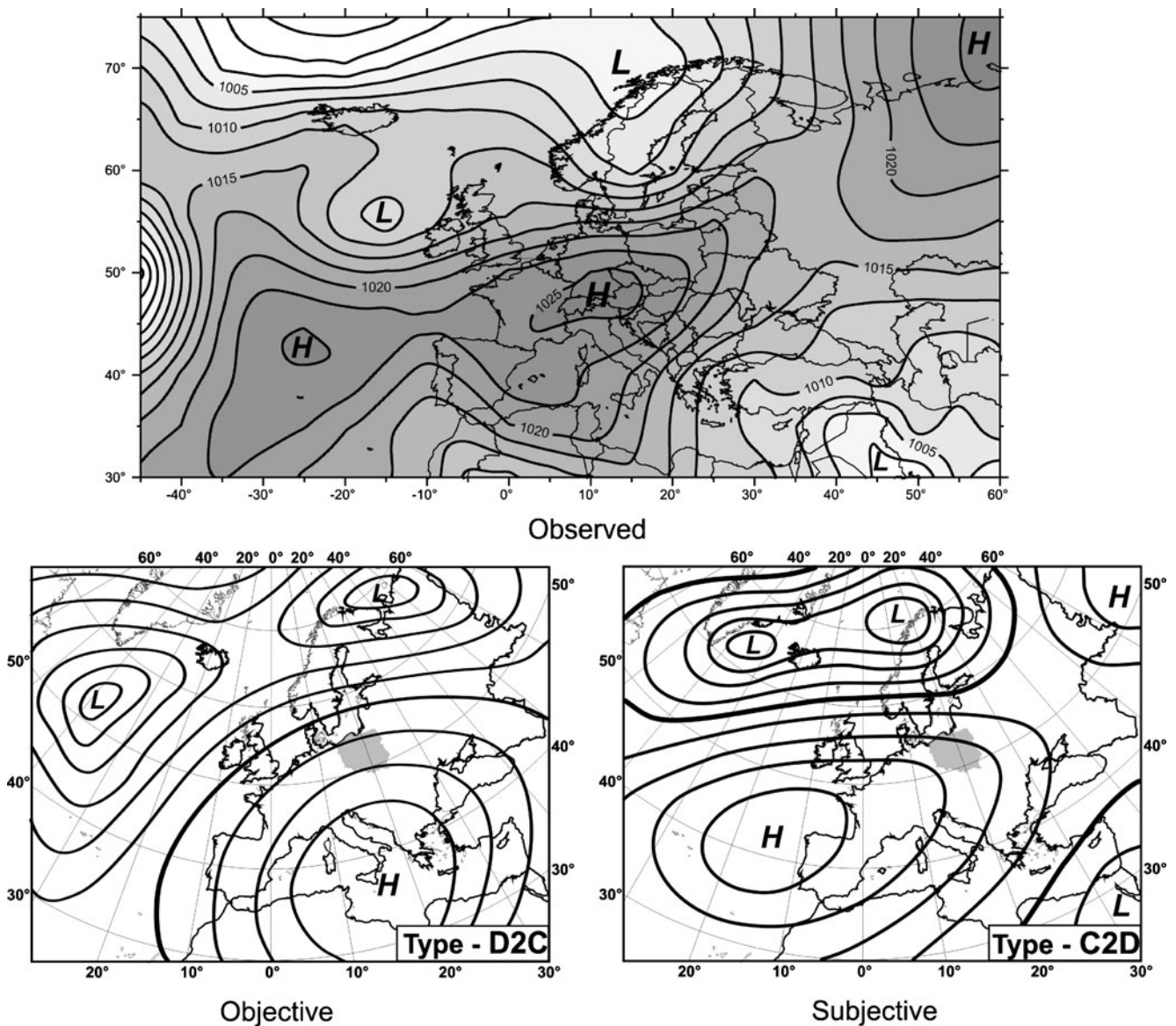
#### 4.2 The second “simple comparison approach”

The analysis of the contingency table produced a very nice surprise, with most of the circulation types described for 1951–1998, using both methods, being identical. This mainly applies to types G, E2C, and E, which were in agreement with types described subjectively in as many as about 50% of the cases (Table 1). Therefore, the challenge was dealing with types whose similarity level was the smallest, ranging from 20% to 30% (types B, D2C, E0). Air-pressure distributions were analyzed in order to find possible reasons for these large differences. The answer, however, turned out to be quite ambiguous. In some cases, the use of grid points was a limitation compared to using a map where all values are continuous, such as in the case of a synoptic map, the type used for subjective evaluation. In other cases, such as on June 30, 1991 (Fig. 3), the air-pressure pattern over Europe was so complex that the proposed control point could not possibly show differences correctly. As a result, subjective and objective circulation types were entirely different from each other, even though they identified two very similar situations (D2C and C2D).

**Table 1** Percentage of cases of types generated via the subjective method and the objective method (100%—number of cases of a particular type based on the objective method)

	o b j e c t i v e													
	TYPE	A	CB	D	B	F	C2D	D2C	G	E2C	E0	E	E1	BE
s u b j e c t i v e	A	38	21	15	3	1	15	14	6	3	5	4	3	5
	CB	9	39	7	12	9	2	2	1	2	6	0	2	4
	D	6	4	27	5	1	1	4	0	3	1	1	2	5
	B	3	2	6	24	4	5	1	0	5	0	0	1	1
	F	0	2	2	9	33	0	0	0	1	10	6	8	18
	C2D	13	9	2	11	2	37	8	4	6	3	1	4	0
	D2C	5	1	3	0	11	6	22	6	2	2	1	1	0
	G	9	2	1	2	10	12	14	52	7	10	14	13	3
	E2C	3	10	2	7	8	6	5	8	46	8	5	2	0
	E0	0	0	1	2	7	1	1	2	3	21	12	5	1
	E	9	8	11	9	3	11	15	14	19	26	54	25	13
	E1	2	2	10	5	8	3	11	6	1	5	2	27	12
	BE	3	0	13	11	3	1	3	1	2	3	0	7	38





**Fig. 3** Situation on June 30, 1991; observed situation and set pattern matched by the subjective/manual and objective methods

It appears that other domains and control points ought to be tested in order to improve the results. However, even a choice of the best possible parameters will not preclude differences, which will be produced via a better objective evaluation of a given circulation type by means of fuzzy logic. This approach will also identify similar types (e.g., E0 and E or D2C and C2D).

Fuzzy logic circulation types calendar based on the Osuchowska-Klein scheme for the entire investigated period 1951–2006 is available on the web site: <http://www.klimat.geo.uj.edu.pl/circulation.htm>. The original Osuchowska-Klein circulation types calendar for the period of 1951–1975 (the only period of the whole accessible series permitted to be published) is also available at the mentioned address.

## 5 Conclusions

Fuzzy logic appears to be quite a promising and effective approach for the identification of circulation types. The results obtained in this study confirmed that fuzzy logic can be used to make traditional (subjective) classification methods objective. The Osuchowska-Klein classification approach has also been useful in this study.

The nature of fuzzy logic allows for the flexible matching of a type with a given pattern. It imitates the way a synoptic meteorologist looks at a synoptic pattern. The fuzzy logic method takes into account the distribution and spatial relationships found across a pressure system. Absolute values of air pressure, though significant, are not

the fundamental matching feature. Therefore, it is always necessary to normalize data before use.

Methods based on fuzzy logic, much like other objective methods, require a long testing period. They allow for the best parameter values to be identified. Nevertheless, even the best set of parameters selected will make the new “objective” classification different from the manual one.

The results of this study have shown that there exists a real problem of an “undefined class” created by the traditional approach. It is equally important to define threshold values, above which the matching of the case and the pattern would be out of the question. Despite the imperfections of the method presented herein, it must be said that its use allows for the definition of types while taking into account areas located between pressure centers, which happens to be very important in the case of Poland. These types of situations are commonly encountered over the area of interest and shape of the weather in the region.

**Acknowledgments** The work was supported by the Polish Ministry of Science and Higher Education under Grant No. COST/2/2006.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Bardossy A, Disse M (1993) Fuzzy rule-based models for infiltration. *Water Resour Res* 29(2):373–382
- Bardossy A, Duckstein L, Bogardi I (1995) Fuzzy rule-based classification of atmospheric circulation patterns. *Int J Climatol* 15:1087–1097
- Cawley GC, Dorling SR (1996) Reproducing a subjective classification scheme for atmospheric circulation patterns over the United Kingdom using a neural network, artificial neural networks. *ICANN* 96:281–286
- Driankov D, Hellendoorn H, Reinfrank M (1996) Introduction to fuzzy logic, (in Polish). WNT, Warszawa, p 320
- Dubois D, Prade H (1980) Fuzzy sets and systems: theory and applications. Academic, New York
- Han J, Kamber M (2001) Data mining: concepts and techniques. Academic, Morgan Kaufmann, San Francisco, p 500
- Hansen BK (2000) Weather prediction using case-based reasoning and fuzzy set theory, master of computer science thesis. Technical University of Nova Scotia, Halifax
- Huth R, Beck Ch, Philipp A, Demuzere M, Ustrnul Z, Cahynova M, Kysely J, Tveito OE (2008) Classifications of atmospheric circulation patterns, recent advances and applications. *Ann NY Sci* 1146:105–152
- James PM (2007) An objective classification method for Hess and Brezowsky Grosswetterlagen over Europe. *Theor Appl Climatol* 88:17–42
- Jenkinson AF, Collison BP (1977) An initial climatology of gales over the North Sea, synoptic climatology branch memorandum, No. 62, Meteorological Office, Bracknell.
- Kendall MG, Stuart A (1973) The advanced theory of statistics, vol 2. Inference and Relationship, Griffin
- Kolodner J (1993) Case-based reasoning. Morgan Kaufmann, San Francisco, p 668
- Leake DB (1996) In: Lake DB (ed) 1996 Case-based reasoning: experiences, lessons & future directions. American Association for Artificial Intelligence, Menlo Park California, pp 3–30
- Osuchowska-Klein B (1975) Forecast aspects of atmospheric circulation over Poland (in Polish, En. summary). *Prace IMGW*, Nr 7, Warszawa, pp 4–51
- Osuchowska-Klein B (1978) Catalogue of atmospheric circulation (in Polish, En. summary). IMGW, Warszawa
- Wilks DS (2006) Statistical methods in the atmospheric sciences, 2nd edn. Academic, San Diego, p 467